NYCU Pattern Recognition, Homework 2

Deadline: April 6, 23:59

Part. 1, Coding (60%):

In this coding assignment, you are required to implement Fisher's linear discriminant by using only NumPy, then train your model on the provided dataset, and evaluate the performance on testing data. Find the sample code and data on the GitHub page https://github.com/NCTU-VRDL/CS_AT0828/tree/main/HW2

Please note that only NumPy can be used to implement your model, you will get 0 point by calling sklearn.discriminant_analysis.LinearDiscriminantAnalysis.

1. (5%) Compute the mean vectors m_i (i=1, 2) of each 2 classes on **training data**

```
mean vector of class 1:
 [[2.47107265]
 [1.97913899]]
mean vector of class 2:
 [[1.82380675]
 [3.03051876]]
```

2. (5%) Compute the within-class scatter matrix S_W on **training data**

```
Within-class scatter matrix SW:
 [[140.40036447  -5.30881553]
 [ -5.30881553 138.14297637]]
```

3. (5%) Compute the between-class scatter matrix S_B on **training data**

```
Between-class scatter matrix SB:
 [[ 0.41895314 -0.68052227]
 [-0.68052227  1.10539942]]
```

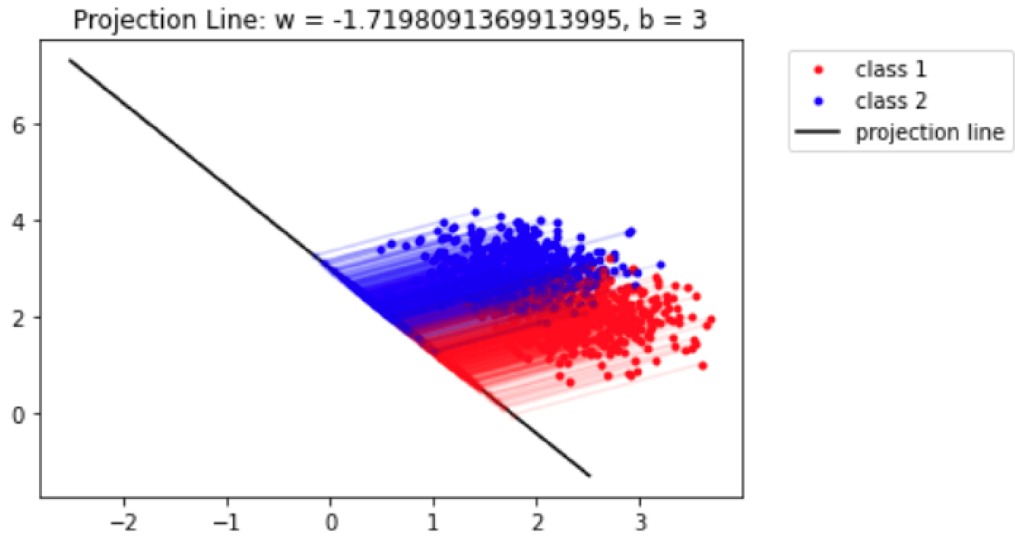
4. (5%) Compute the Fisher's linear discriminant W on **training data**

```
Fisher's linear discriminant:
 [[ 0.50266214]
 [-0.86448295]]
```

5. (20%) Project the **testing data** by Fisher's linear discriminant to get the class prediction by nearest-neighbor rule and calculate your accuracy score on **testing data** (you should get accuracy over 0.9)

```
Accuracy of test-set: 0.9
```

6. (20%) Plot the **1) best projection line** on the **training data** and show the slope and intercept on the title *(you can choose any value of* ***intercept*** *for better visualization)* **2) colorize the data** with each class **3)** project all data points on your projection line.



Part. 2, Questions (40%):

1. (10%) Show that maximization of the class separation criterion given by $L(\lambda, w) = w^T(m2 - m1) + \lambda(w^T w - 1)$ with respect to w, using a Lagrange multiplier to enforce the constraint $w^T w = 1$, leads to the result that $w \propto (m2 - m1)$.

$L = w^T(m_2 - m_1) + \lambda(w^T w - 1)$

taking the gradient of L we obtain $\nabla L = m_2 - m_1 + 2\lambda w$

and setting this gradient to zero gives $w = -\frac{1}{2\lambda}(m_2 - m_1)$

from which it follows that $w \propto (m_2 - m_1)$

2. (15%) By making use of (eq 1), (eq 2), (eq 3), (eq 4), and (eq 5), show that the Fisher criterion (eq 6) can be written in the form (eq 7).

$$(m_2-m_1)^2 = [w^T(m_2-m_1)]^2 = w^T(m_2-m_1)(m_2-m_1)^T w = w^T S_B w$$

$$s_1^2 = \sum_{n\in C_1} (y_n-m_1)^2 = \sum_{n\in C_1} (w^T x_n - w^T m_1)^2$$

$$= \sum_{n\in C_1} w^T (x_n-m_1)(x_n-m_1)^T w$$

$$= w^T S_1 w$$

同理， $s_2^2 = w^T S_2 w$

$$\therefore s_1^2+s_2^2 = w^T S_1 w + w^T S_2 w = w^T (S_1+S_2) w = w^T S_W w$$

$$\therefore J(w) = \frac{(m_2-m_1)^2}{s_1^2+s_2^2} = \frac{w^T S_B w}{w^T S_W w}$$

3. (15%) By making use of the result (eq 8) for the derivative of the logistic sigmoid, show that the derivative of the error function (eq 9) for the logistic regression model is given by (eq 10), where $y_n = \sigma(a_n)$, $a_n = w^T\varphi_n$.

$\because y_n = \sigma(w^T\phi_n)$

$$\nabla_w E(w) = -\sum_{n=1}^{N}\left[\frac{t_n}{y_n}y_n' + \frac{1-t_n}{1-y_n}(-y_n')\right], \quad \text{where } y_n' = y_n(1-y_n)\phi_n$$

$$\Rightarrow \nabla_w E(w) = -\sum_{n=1}^{N}\left[\frac{t_n}{y_n}y_n(1-y_n)\phi_n - \frac{1-t_n}{1-y_n}y_n(1-y_n)\phi_n\right]$$

$$= \sum_{n=1}^{N}(y_n - t_n)\phi_n$$

(eq 1) $$y = \mathbf{w}^{\mathrm{T}}\mathbf{x}.$$

(eq 2) $$m_2 - m_1 = \mathbf{w}^{\mathrm{T}}(\mathbf{m}_2 - \mathbf{m}_1)$$

(eq 3) $$s_k^2 = \sum_{n \in \mathcal{C}_k} (y_n - m_k)^2$$

(eq 4) $$\mathbf{S}_{\mathrm{B}} = (\mathbf{m}_2 - \mathbf{m}_1)(\mathbf{m}_2 - \mathbf{m}_1)^{\mathrm{T}}$$

(eq 5) $$\mathbf{S}_{\mathrm{W}} = \sum_{n \in \mathcal{C}_1} (\mathbf{x}_n - \mathbf{m}_1)(\mathbf{x}_n - \mathbf{m}_1)^{\mathrm{T}} + \sum_{n \in \mathcal{C}_2} (\mathbf{x}_n - \mathbf{m}_2)(\mathbf{x}_n - \mathbf{m}_2)^{\mathrm{T}}$$

(eq 6) $$J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$$

(eq 7) $$J(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{B}}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{W}}\mathbf{w}}$$

(eq 8) $$\frac{d\sigma}{da} = \sigma(1 - \sigma).$$

(eq 9) $$E(\mathbf{w}) = -\ln p(\mathbf{t}|\mathbf{w}) = -\sum_{n=1}^{N} \{t_n \ln y_n + (1 - t_n)\ln(1 - y_n)\}$$

(eq 10) $$\nabla E(\mathbf{w}) = \sum_{n=1}^{N} (y_n - t_n)\phi_n$$